

Recherche de variables environnementales explicatives de la qualité des graines de colza

C.Lanoix and C.Pontet

Terres Inovia

Centre de recherche INRA, Chemin de Borde-Rouge, 31326 Castanet-Tolosan
c.lanoix@terresinovia.fr

Mots clefs : multicollinéarité, régression PLS, régression Lasso, colza, qualité des graines

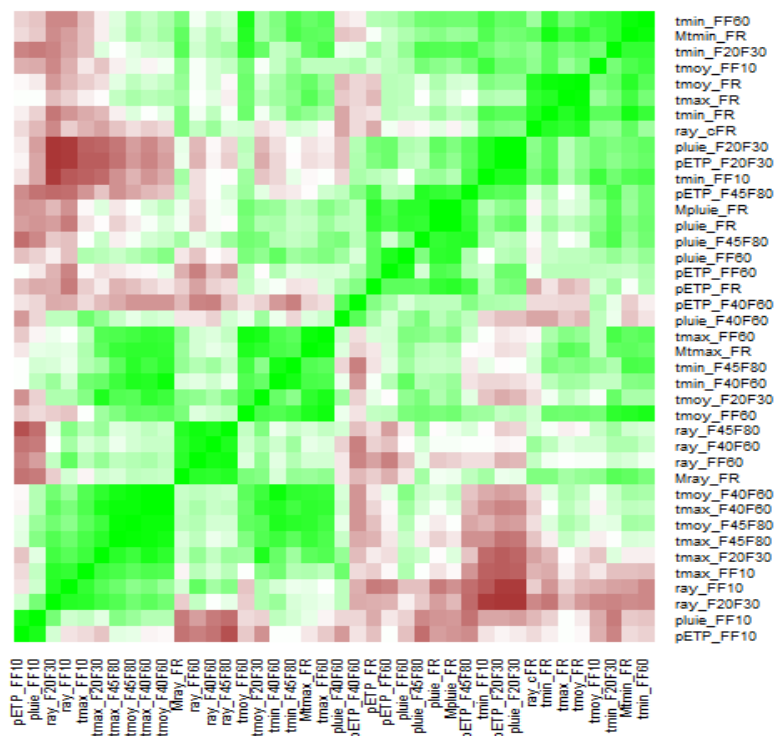
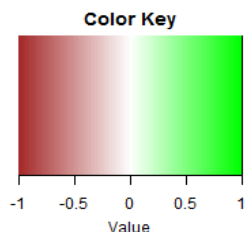
Avec plus de 6 millions d'hectares et 20 millions de tonnes de production par an, le colza est la première culture oléagineuse Européenne. Un de ses principaux débouchés est l'alimentation humaine, avec la possibilité de produire une huile de haute qualité, riche en acides gras omega3.

Par ailleurs les tourteaux, coproduits de la trituration des graines de colza (extraction de l'huile), constituent une source de protéines de qualité pour les bovins, ovins et porcins. Ces produits issus des graines de colza représentent des productions à haute valeur ajoutée, soumises à des cahiers des charges précis. Or, le contexte actuel de changement climatique et de réduction des intrants, conduit à une diversification des milieux de cultures, avec une plus grande expression des stress.

Afin de remplir les cahiers des charges qui leur sont imposés, les agriculteurs doivent alors se placer dans des conditions favorables en adaptant le choix de la variété au milieu de culture. Cela implique de comprendre et prédire les interactions génotype x environnement x conduite (IGEC) [1] s'appliquant sur la qualité de la graine. L'objectif de cette étude, réalisé dans le cadre du projet POLYGONE¹, est alors de rechercher un ensemble de variables environnementales explicatives de la qualité des graines.

Les données utilisées sont issues d'un réseau de 10 lieux sur lesquels 8 variétés de colza ont été cultivées. Pour chaque individu, autrement dit chaque variété dans un lieu, la qualité des graines est définie par les teneurs en huile, protéines et glucosinolates, ainsi que par leur profil en acide gras, dont particulièrement les acides oléique (C18:1 ω -9), linoléique (C18:2 ω -6) et linoléique (C18:3 ω -3) [2]. Chaque lieu est caractérisé par un ensemble de variables candidates calculées à partir de données météorologiques, quantifiant les stress rencontrés, comme le déficit hydrique [3], les fortes ou très basses températures [4], et ceci sur les stades sensibles de culture [5].

L'ensemble des 40 variables explicatives candidates, sont très corrélées entre elles (cf. figure ci-dessous). Les méthodes statistiques linéaires sont donc à proscrire dans un premier temps. Nous ferons donc appel à des méthodes telles que la régression PLS [6] et LASSO [7] [8]. Ces approches nous permettront de restreindre par régularisation l'espace des solutions, afin de limiter l'effet de la multicollinéarité des covariables. Aussi, divers critères comme R^2 ajusté, PRESS, RMSE, VIP seront utilisés afin d'évaluer la qualité explicative de chaque variables, pour ainsi sélectionner les plus pertinentes.



Dans le but d'améliorer le conseil variétal, cette étude représente l'étape préliminaire indispensable à l'analyse des interactions génotype x environnement x conduite, appliquée à la qualité des graines de colza.

¹ Ce travail a été réalisé, en partenariat avec la SAS PIVERT, dans le cadre de l'Institut pour la Transition Energétique (ITE) P.I.V.E.R.T. (www.institut-pivert.com) retenu parmi les Investissements d'Avenir. Ce travail a bénéficié d'une aide de l'Etat au titre du Programme d'Investissements d'Avenir portant la référence ANR-001-01.

Références

[1] P.Debaeke, P.Casadebaig, E.Mestries, J.P.Palleau, F.Salvi, V.Bertoux, V.Uyttewaal. (2011) Evaluer et valoriser les interactions variété-milieu-conduite en tournesol. *Innovations Agronomique* 14, 77-90

[2] Terres Inovia, (2015). Fiche : qualité des graines de colza, récolte 2015

[3] L.Champolivier, A.Merrien. (1996). Effects of water stress applied at different growth stages to *Brassica napus* L. var. *oleifera* on yield, yield components and seed quality. *European Journal of Agronomy* 5, 153-160.

[4] A.Baux, T.Hebeisen, D.Pellet. Effect of temperature on low-linolenic rapeseed oil fatty-acid composition. *AGRONOMY : Farming Systems and Ecology*

[5] L.Champolivier, Dossier Oléoscope n°86 (Septembre 2006). Le remplissage des graines, étape cruciale pour l'élaboration du rendement et la qualité du colza.

[6] Tenenhaus, M. (1998). La régression PLS, théorie et pratique. Editions TECHNIP.

[7] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267-288.

[8] G.James, D.Witten, T.Hastie, R.Tibshirani. (2013). An introduction to statistical learning with applications in R, Springer